

PATENT APPLICATION

**TECHNIQUES FOR USING PAPER DOCUMENTS AS MEDIA
TEMPLATES**

Inventor(s): Berna Erol, a citizen of Canada, residing at
20380 Stevens Creek Blvd, Unit #326
Cupertino, CA 95014

Jonathan J. Hull, a citizen of The United States, residing at
751 Laurel Street, PMB 434
San Carlos, CA 94070

Jamey Graham, a citizen of The United States, residing at
1196 Shasta Avenue
San Jose, CA 95126

Assignee: Ricoh Company, Ltd.
15-5 Minami Aoyama 1-Chome, Minato-ku
Tokyo, JAPAN 107-8544

Entity: Large

TECHNIQUES FOR USING PAPER DOCUMENTS AS MEDIA TEMPLATES

CROSS-REFERENCES TO RELATED APPLICATIONS

- 5 **[0001]** The present application incorporates by reference for all purposes the entire contents of the following:
- [0002]** U.S. Application No. 09/728,560, entitled "TECHNIQUES FOR CAPTURING INFORMATION DURING MULTIMEDIA PRESENTATIONS", filed November 30, 2000 (Attorney Docket No. Docket No. 15358-006210US);
- 10 **[0003]** U.S. Application No. 09/728,453, entitled "TECHNIQUES FOR RECEIVING INFORMATION DURING MULTIMEDIA PRESENTATIONS & COMMUNICATING THE INFORMATION", filed November 30, 2000 (Attorney Docket No. Docket No. 15358-006220US);
- [0004]** U.S. Application No. 09/521,252, entitled "METHOD & SYSTEM FOR
- 15 INFORMATION MANAGEMENT TO FACILITATE THE EXCHANGE OF IDEAS DURING A COLLABORATIVE EFFORT", filed March 8, 2000 (Attorney Docket No. 15358-006300US);
- [0005]** U.S. Application No. 10/001,895, entitled "PAPER-BASED INTERFACE FOR MULTIMEDIA INFORMATION", filed November 19, 2001 (Attorney Docket No. Docket
- 20 No. 15358-006500US);
- [0006]** U.S. Application No. 10/660,985, entitled "TECHNIQUES FOR STORING MULTIMEDIA INFORMATION WITH SOURCE DOCUMENTS", filed September 12, 2003 (Attorney Docket No. 15358-009410US);
- [0007]** U.S. Application No. 10/661,052, entitled "TECHNIQUES FOR PERFORMING
- 25 OPERATIONS ON A SOURCE SYMBOLIC DOCUMENT", filed September 12, 2003 (Attorney Docket No. 15358-009420US);
- [0008]** U.S. Application No. 10/660,867, entitled "TECHNIQUES FOR ACCESSING INFORMATION CAPTURED DURING A PRESENTATION USING A PAPER

DOCUMENT FOR THE PRESENTATION", filed September 12, 2003 (Attorney Docket No. 15358-009430US);

[0009] U.S. Application No. 10/696,735, entitled " TECHNIQUES FOR USING A CAPTURED ELECTRONIC REPRESENTATION FOR THE RETRIEVAL OF
5 RECORDED INFORMATION", filed September 12, 2003 (Attorney Docket No. 15358-009440US); and

[0010] U.S. Application No. 10/412,757, entitled "AUTOMATED TECHNIQUES FOR COMPARING CONTENTS OF IMAGES", filed April 11, 2003 (Attorney Docket No. Docket No. 15358-009500US).

10

BACKGROUND OF THE INVENTION

[0011] The present application relates to field of accessing recorded information, and more particularly to techniques for creating an electronic representation that includes inserted information that is related to recorded information.

15 [0012] Recording information during presentations has gained a lot of popularity in recent years. For example, colleges and universities have started to program classes and lectures, corporations have started to record meetings and conferences, etc. One or more capture devices may record information during a presentation. The recorded information may comprise different types or streams of information including audio information, video
20 information, and the like.

[0013] After the presentation, the recorded information is then available for use by a user. A user may review their notes and may want to view the recording of the presentation. The conventional way for accessing these recordings has been by viewing the recordings sequentially. More efficient techniques are desired for accessing or retrieving the recorded
25 information or indexing into the recorded information.

BRIEF SUMMARY OF THE INVENTION

[0014] Embodiments of the present invention generally relate to techniques for creating a composite electronic representation. The techniques include receiving an electronic
30 representation of a paper document. Features in the electronic representation are then extracted and compared to recorded information to determine matching information. For

example, the matching information may be presentations and/or pages in the recorded information. Information to insert is then determined based on the matching information and the received electronic representation. The composite electronic representation is then created using the determined information.

5 **[0015]** In one embodiment, a method for creating a composite electronic representation is provided. The method comprises: receiving an electronic representation of a document; extracting a feature from the electronic representation of the document; comparing the feature to the recorded information to determine information in the recorded information that matches the feature; determining information to insert based on the information in the
10 recorded information that matches the feature and the received electronic representation of a document; and creating a composite electronic representation comprising the determined information.

[0016] In another embodiment, a method for creating a composite electronic representation of a document using information recorded during a presentation is provided. The method
15 comprises: receiving an electronic representation of a document for the presentation, the electronic representation including a feature that was presented during the presentation; extracting the feature from the electronic representation; comparing the feature to the information recorded during the presentation to determine information in the recorded information that matches the one or more features; and determining information to insert
20 based on the information in the recorded information that matches the feature and the received electronic representation of a document; and creating a composite electronic representation comprising the determined information.

[0017] The foregoing, together with other features, embodiments, and advantages of the present invention, will become more apparent when referring to the following specification,
25 claims, and accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] Fig. 1 is a simplified block diagram of a system that may incorporate an embodiment of the present invention;

30 **[0019]** Fig. 2 depicts a simplified flowchart of a method for using an electronic representation of a paper document to create an electronic representation with inserted

information related to recorded information according to one embodiment of the present invention;

[0020] Fig. 3A depicts a page of a paper document according to one embodiment of the present invention;

5 [0021] Fig. 3B depicts the page of Fig. 3A with user markings according to one embodiment of the present invention;

[0022] Fig. 3C depicts an interface that includes information inserted that is related to recorded information according to one embodiment of the present invention;

10 [0023] Fig. 4 depicts a possible output after information is selected according to one embodiment of the present invention;

[0024] Fig. 5 is a simplified block diagram of modules that may be used to implement an embodiment of the present invention;

[0025] Fig 6 depicts paper document templates according to embodiments of the present invention;

15 [0026] Fig. 7 depicts keyframes according to one embodiment of the present invention; and

[0027] Fig. 8 is a simplified block diagram of data processing system that may be used to perform processing according to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

20 [0028] In the following description, for the purposes of explanation, specific details are set forth in order to provide a thorough understanding of the invention. However, it will be apparent that the invention may be practiced without these specific details.

[0029] Fig. 1 is a simplified block diagram of a system 100 that may incorporate an embodiment of the present invention. System 100 depicted in Fig. 1 is merely illustrative of
25 an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[0030] System 100 includes a computer system 102 that may be used by a user to prepare material to be presented at a presentation. Examples of presentations include lectures,

meetings, conferences, classes, speeches, demonstrations, etc. The presentation material may include slides, photos, audio messages, video clips, text information, web pages, etc. The user may use one or more applications 104 executed by computer 102 to generate the presentation material. An example of a commonly used application for preparing slides to be presented at a presentation is PowerPoint™ provided by Microsoft™ Corporation. For example, as depicted in Fig. 1, the user may use PowerPoint™ application 104 to create a "presentation.ppt" file 106 (*.ppt file). A *.ppt file created using a PowerPoint™ application can comprise one or more pages, each page comprising one or more slides. A *.ppt file may also store information as to the order in which the slides are to be presented at the presentation and the manner in which the slides will be presented.

[0031] In addition to PowerPoint™ presentation files comprising slides, other types of files comprising other presentation material may also be created using different applications executed by computer 102. These files may be referred to in general as "symbolic presentation files". A symbolic presentation file is any file created using an application or program and that comprises at least some content that is to be presented or output during a presentation. A symbolic presentation file may comprise various types of contents such as slides, photos, audio messages, video clips, text, web pages, images, etc. A *.ppt file created using a PowerPoint™ application is an example of a symbolic presentation file that comprises slides.

[0032] The user may print portions of the presentation material on a paper medium to generate paper documents (also referred to as "paper documents") that are usually handed out at the presentation. The term "paper medium" is intended to refer to any tangible medium on which information can be printed. The term "print" or "printing" is intended to include writing, imprinting, drawing, embossing, and the like. Each paper document may comprise one or more paper pages. Depending on the number of people attending the presentation, multiple paper documents may be printed.

[0033] An electronic representation of a paper document is received. As shown in Fig. 1, scanner 108 may be used to scan a paper document 110. Various other devices that are capable of scanning information on a paper medium may also be used to scan paper documents. Examples of such devices include facsimile machines, copiers, scanners, and the like.

[0034] Various different features may be on a paper document. Generally, the features on a document relate to information to be presented or discussed during the presentation for which the document is created. The features may include portions of presentation material or other material. Examples of features that can be printed include slides, photos, web pages, text
5 information (e.g., a list of agenda features to be discussed at a meeting), and the like. For example, the user may print one or more slides from a *.ppt file on a paper document. The PowerPoint™ application provides tools for printing one or more slides from a *.ppt file to generate a paper document. Each page of the paper document may have one or more slides printed on it. Examples of paper document pages with slides on them are depicted in Figs.
10 3A, and 3B and described below in further detail.

[0035] The electronic representation may also be an electronic image of paper document. For example, a *.ppt file may be converted to images and used as the electronic representation of the paper document. Although paper document is used, an electronic representation of any document may be received. A document that is printed in paper does
15 not need to be used to generate the electronic representation.

[0036] Capture devices 118 are configured to capture information presented at a presentation. Various different types of information output during a presentation may be captured or recorded by capture devices 118 including audio information, video information, images of slides or photos, whiteboard information, text information, and the like. For
20 purposes of this application, the term "presented" is intended to include displayed, output, spoken, etc. For purposes of this application, the term "capture device" is intended to refer to any device, system, apparatus, or application that is configured to capture or record information of one or more types. Examples of capture devices 118 include microphones, video cameras, cameras (both digital and analog), scanners, presentation recorders, screen
25 capture devices (e.g., a whiteboard information capture device), symbolic information capture devices, etc. In addition to capturing the information, capture devices 118 may also be able to capture temporal information associated with the captured information.

[0037] A presentation recorder is a device that is able to capture information presented during a presentation, for example, by tapping into and capturing streams of information from
30 an information source. For example, if a computer executing a PowerPoint™ application is used to display slides from a *.ppt file, a presentation recorder may be configured to tap into the video output of the computer and capture keyframes every time a significant difference is

detected between displayed video keyframes of the slides. The presentation recorder is also able to capture other types of information such as audio information, video information, slides information stream, etc. The temporal information associated with the captured information indicating when the information was output or captured is then used to
5 synchronize the different types of captured information. Examples of presentation recorders include a screen capture software application, a PowerPoint™ application that allows recording of slides and time elapsed for each slide during a presentation, presentation recorder described in U.S. Application No. 09/728,560, filed November 30, 2000 (Attorney Docket No. 15358-006210US), U.S. Application No. 09/728,453, filed November 30, 2000
10 (Attorney Docket No. 15358-006220US), and U.S. Application No. 09/521,252, filed March 8, 2000 (Attorney Docket No. 15358-006300US), and are hereby incorporated by reference for all purposes.

[0038] A symbolic information capture device is able to capture information stored in symbolic presentation documents that may be output during a presentation. For example, a
15 symbolic information capture device is able to record slides presented at a presentation as a sequence of images (e.g., as JPEGs, BMPs, etc.). A symbolic information capture device may also be configured to extract the text content of the slides. For example, during a PowerPoint™ slide presentation, a symbolic information capture device may record the slides by capturing slide transitions (e.g., by capturing keyboard commands) and then extracting the
20 presentation images based on these transitions. Whiteboard capture devices may include devices such as a camera appropriately positioned to capture contents of the whiteboard, a screen, a chart, etc.

[0039] The information captured or recorded by capture devices 118 during a presentation may be stored in a repository or database 115 as recorded information 120. Recorded
25 information 120 may be stored in various formats. For example, a directory may be created in repository 115 for storing recorded information 120, and the various types of information (e.g., audio information, video information, images, etc.) included in recorded information 120 may be stored in the directory. In another embodiment, recorded information 120 may be stored as a file. Various other techniques known to those skilled in the art may also be
30 used for storing the recorded information.

[0040] Images of the slides found in the paper document are displayed during a presentation. In one embodiment, a presentation recorder may capture slide images as they

are displayed. In addition, association information that may be used to index into recorded information 120 may be stored. For example, time information may be stored indicating a time that the slide was displayed. The time information may then be used to determine portions of recorded information 120 that correspond to when the slide was displayed.

5 [0041] In addition to the time information, source information identifying the location where the recorded information for the presentation is stored may also be determined. This storage location information for recorded information 120 may be updated when the recorded information is moved to a new storage location. In this manner, embodiments of the present invention allow the storage location of recorded information 120 to be changed over time.

10 [0042] According to an embodiment of the present invention, the association information is stored in the XML structure. For example, the association information may include time information and source information for presentations and/or pages determined in step 206. The source information may be an identifier used to access a presentation. For example, the source information may be a location and file name. The time information is then used to
15 index into a portion of a presentation. The presentation may have been a presentation determined in step 204 or a presentation that is related to information determined in step 204 (e.g., a presentation from which slide images were captured using a presentation recorder). The portion of the presentation may include information that matches the extracted features determined in step 204. For example, the portion of the presentation may include a slide that
20 was displayed during the presentation.

[0043] Server 112 creates a composite electronic representation 122. In one embodiment, features are extracted from an electronic representation of a paper document received. The features are compared to recorded information 120 to determine matching information. In one embodiment, matching information may be determined using various techniques. It will
25 be recognized that a slide image may not have to match a slide image exactly. For example, text in a slide may be compared with text to determine text that substantially matches.

[0044] Information to insert is then determined based on the matching information and the electronic representation of the paper document. Composite electronic representation 122 is then created based on the information inserted. Composite electronic representation 122 may
30 include the features extracted. Also, composite electronic representation 122 may include the information determined based on the matching information and the electronic representation of the paper document.

[0045] A user may select the inserted information in composite electronic representation 122 and have recorded information 120 accessed and/or played. For example, an electronic representation of a slide in the paper document may be used to determine information that is related to recorded information 120. The inserted information may include an object showing
5 a picture of a portion of a presentation. When the object is selected, recorded information 120 when the slide was displayed during the presentation is accessed and/or played. Thus, when a user desires additional information related to a slide in the paper document, the inserted information in composite electronic representation 122 may be used to retrieve recorded information 120 of the presentation when the slide was displayed and/or discussed.

10 [0046] Fig. 2 depicts a simplified flowchart 200 of a method for a composite electronic representation 122 according to one embodiment of the present invention. The method depicted in Fig. 2 may be performed by software modules executed by a processor, hardware modules, or combinations thereof. Flowchart 200 depicted in Fig. 2 is merely illustrative of an embodiment of the present invention and is not intended to limit the scope of the present
15 invention. Other variations, modifications, and alternatives are also within the scope of the present invention.

[0047] In step 202, an electronic representation of a document is received. In one embodiment, the document was printed as a paper document. The user may have taken notes on the paper document. The notes may be typically written on a paper document for a
20 presentation. Other ways of taking notes may be appreciated also, such as typing in notes on an electronic version of the document. Although notes may be written on the paper document, it is not necessary that any notes be present.

[0048] Although an electronic representation of a paper document will be assumed as received, it will be understood that an electronic copy of a document may be received. One
25 or more pages in the paper document may be received as the electronic representation. For discussion purposes, it is assumed that the electronic representation includes a single page but it should be understood that the electronic representation may include any number of pages. As discussed above, an electronic representation of a paper document may be received from a document that is scanned.

30 [0049] In another embodiment, an electronic version of the paper document may be used. For example, the electronic version may be images of slides found in a *.ppt file. A *.ppt file may be converted to images. For example, techniques are known where *.ppt slides may be

converted into images in a .pdf or flash file. These images may be used as images of a document. For example, an electronic representation of a document may be received using the following process. A user may open the electronic representation in an application, such as a pdf reader. For example, the electronic representation from the scanned document or the
5 electronic version may be opened in the application. An input may be provided in the pdf reader that initiates the following steps.

[0050] In step 204, features are extracted from the electronic representation of the document. The document may include images of one or more slides that were presented in the presentation. The images of the slides are then extracted from the electronic
10 representation of the document. Although the processing is described as extracting slides, it will be understood that other features may be extracted. For example, pictures, text, etc. may be extracted. For example, instead of using slides in a presentation, the user may display pictures of certain feature s. The pictures are then extracted.

[0051] In one embodiment, segmentation may be used if more than one slide is present on a
15 page. Segmentation separates and determines separate slides on the electronic representation of the paper document. This may be desirable when the individual slides should be associated with different portions of recorded information 120. Segmentation may not be required in some cases, such as when only one slide is on a page, etc.

[0052] Techniques for segmenting documents and images are well known in the art. Many
20 of the current segmentation techniques may be used for segmenting individual slide regions from an electronic representation of a paper document. An example of segmentation techniques will be described below in more detail. Although segmentation is described, it should be understood that other techniques may be used to determine slide images in an electronic representation of a paper document. For example, the content of an electronic
25 representation may be analyzed to determine slide images. In one example, a rectangular box may be recognized and the information in the box may be used as a slide image.

[0053] In step 206, extracted features are compared with recorded information 120 to
determine matching information in recorded information 120. For example, portions of recorded information 120 that match the extracted features may be determined. The portions
30 may include, pages of a presentation, video, audio, etc. In one embodiment, slide images extracted from the electronic representation of the document are compared to slide images in recorded information 120 for different presentations. Recorded information 120 for the

presentations may have been presented at different times and thus all include matching information (e.g., one slide may occur in different recorded information for presentations).

[0054] In one embodiment, the extracted features may include a plurality of slides. The slides may be compared to slide images to determine a presentation that includes matching information. In one embodiment, each slide in the plurality of slides is compared to slides in recorded information 120 for presentations. Portions of recorded information 120 that include information that match the slides are determined.

[0055] In another embodiment, the plurality of slides are treated as a set and all slides are used in comparing slides in all presentations. Thus, in order for a presentation to be identified to include matching information, the presentation should include slides that match each of the plurality of slides taken as a set.

[0056] Various techniques may be used to determine matching information in recorded information 120. In one embodiment, techniques described in U.S. Application No. 10/412,757, entitled "AUTOMATED TECHNIQUES FOR COMPARING CONTENTS OF IMAGES" (Attorney Docket No. Docket No. 15358-009500US), filed April 11, 2003; U.S. Application No. 10/660,985, entitled "TECHNIQUES FOR STORING MULTIMEDIA INFORMATION WITH SOURCE DOCUMENTS" (Attorney Docket No. 15358-009410US), filed September 12, 2003; U.S. Application No. 10/661,052, entitled "TECHNIQUES FOR PERFORMING OPERATIONS ON A SOURCE SYMBOLIC DOCUMENT" (Attorney Docket No. 15358-009420US), filed September 12, 2003; U.S. Application No. 10/660,867, entitled "TECHNIQUES FOR ACCESSING INFORMATION CAPTURED DURING A PRESENTATION USING A PAPER DOCUMENT FOR THE PRESENTATION" (Attorney Docket No. 15358-009430US), filed September 12, 2003; U.S. Application No. 10/696,735, entitled "TECHNIQUES FOR USING A CAPTURED ELECTRONIC REPRESENTATION FOR THE RETRIEVAL OF RECORDED INFORMATION" (Attorney Docket No. 15358-009440US), filed September 12, 2003 and other techniques known to those skilled in the art may be used to find matching images (i.e., images from the recorded information that comprise the extracted features).

[0057] In one embodiment, the extracted features may be used to determine recorded information 120 for presentations that include information that match the extracted features. For example, this is done by first extracting images from recorded information 120. The images that are extracted from recorded information 120 may include images captured during

the presentation by various electronic representation capture devices, images captured by a presentation recorder, keyframe images obtained from video information captured during the presentation, and the like.

[0058] The extracted images are then compared with the extracted features determined in step 204. The extracted images may have been pre-processed to determine time information indicating the time(s) during the presentation when the slide was displayed or presented. The time information for a slide may also identify one or more spans of time during the presentation when the slide was presented or displayed. The spans of time may be non-contiguous.

[0059] In one embodiment, the matching information may be determined using techniques described in the "Matching Techniques" section described below. The matching information may be determined using presentation level matching. A document-matching algorithm receives a document electronic representation I_i as input and it compares that to the database of Presentation Recorder documents. This may be referred to as a presentation-matching step. It locates every presentation recording session that may have been used to give the presentation. The next step, called slide matching, maps each segmented slide electronic representation $p_{j,k}$ in the identified Presentation Recorder sessions onto the slide images on the document. The techniques are described in more detail below. This provides the mapping from each slide in the documents to the source and time stamps in the audio and video tracks.

[0060] In step 207, information to insert is determined based on the matching information determined in step 206 and the electronic representation of the paper document received in step 202. In one example, the information determined may be based on slide images that match the extracted features. The association information determined for a matching slide image may be used to index into a presentation. An image extracted from the presentation recording at the time indicated by the association information may then be determined.

[0061] In step 208, a composite electronic representation 122 is created with the information determined in step 207. Composite electronic representation may include many types of information. For example, the features extracted may be included in composite electronic representation 122.

[0062] Also, the information may include metadata that is determined based on recorded information 120. The metadata may be derived from the matching information determined in

step 206. For example, metadata may also be determined and inserted in the created electronic representation. Recorded information 120 may be post processed to extract metadata to be included in composite electronic representation 122 (created using documents as templates). For example, how long a slide was discussed may be calculated and inserted.

5 It should be understood that there is no limit on what kind of metadata is extracted or in what form they can be included in composite electronic representation 122, some examples for extraction of metadata are provided for illustrative purposes. Techniques for determining metadata will be described in more detail below.

[0063] Also, a selectable object, such as an icon or a link, may be inserted. When selected,
10 the object uses association information that is related to the matching information determined in step 206 to access recorded information 120. For example, the accessed information may be a presentation recording at a time when a slide was displayed. Thus, an image extracted from recorded information 120, when selected, may cause the presentation recorded to be accessed at a time specified by association information for the image. In another
15 embodiment, recorded information 120 may be embedded or stored with the image. When an object is selected, the embedded information is accessed and played. Thus, a central database may not need to be accessed. For example, a video player object may be embedded in the image. When play is selected, recorded information 120 is automatically played.

[0064] In one embodiment, the electronic representation received in step 202 is used to
20 create composite electronic representation 122. For example, information is inserted into the electronic representation received. Also, a new document may be created that includes the electronic representation received and the inserted information. In both cases, composite electronic representation 122 is created with information relating to recorded information 120 inserted into it. For example, a paper document may be printed and handed out for a
25 presentation. A user may take notes on the paper document. The paper document is then scanned to generate an electronic representation of the paper document. Information is then inserted in the scanned electronic representation.

[0065] In another embodiment, a document that is different from the electronic
representation received is created. A different document may include the features extracted
30 in step 204 and the inserted information. For example, just the extracted images of slides in a paper document and the determined information may be included in composite electronic representation 122.

[0066] Composite electronic representation 122 created in step 210 may then be emailed, reprinted, stored on a server for later access, copied to storage medium, such as a CD, display, etc. When a user needs to review that particular presentation, the user can review the notes taken in composite electronic representation 122 (assuming that composite electronic representation 122 included the notes of the paper document). If more information is needed, the information inserted in composite electronic representation 122 may be selected and recorded information 120 corresponding to association information for the inserted information may be accessed and displayed. For example, a presentation playback interface may be invoked and playback starts from a time stored in the association information. In other embodiments, the inserted information may be used to add information to composite electronic representation 122. For example, metadata may indicate how long the slide was discussed.

[0067] Fig. 3A depicts a simplified page 300 from a paper document generated according to embodiments of the present invention. Page 300 depicted in Fig. 3 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[0068] As shown in Fig. 3A, information 302 identifying the presentation and the presenter is printed on page 300. Other information such as the time when the presentation takes place, the duration of the presentation, etc. may also be included in information 302. In the embodiment depicted in Fig. 3A, three slides (features) 304-1, 304-2, and 304-3 are printed on page 300. Additionally, spaces 308 are provided for a user to take notes during the presentation for each slide.

[0069] Fig. 3B depicts page 300 of Fig. 3A with user markings according to embodiments of the present invention. As shown, a user has written notes on the document in spaces 308-1 and 308-2.

[0070] Fig. 3C depicts a composite electronic representation 122 according to one embodiment of the present invention. Composite electronic representation 122 may be viewed using an interface, such as a pdf reader, web browser, word processing interface, etc. Composite electronic representation 122 includes information inserted that is related to recorded information 120 according to one embodiment of the present invention. As shown in interface 310, composite electronic representation 122 that includes at least a portion of

page 300. Composite image in Fig. 3C may be created using page 300 as a base. Other information may be superimposed on composite image.

[0071] Composite electronic representation 122 also includes information 314. As shown, images of recorded information 120 are included in information 314. The images, in one embodiment, correspond to a portion in recorded information 120 for a presentation. The portion may be when a slide that matches a slide 304 was outputted. For example, information 314-1 includes information extracted from recorded information 120 where an image of slide 304-1 was outputted.

[0072] In one embodiment, information 314, when selected, may cause an action to be performed. Each image in information 314 may be associated with association information, such as time and source information, that is used to access recorded information 120. Although not shown, information 314 may also include information other than images, such as hypertext links, icons, metadata, etc.

[0073] Composite electronic representation 122 may include the received electronic representation of paper document 300. In this case, information 314 is inserted into the scanned electronic representation. Thus, a user that took notes on a paper document may view the paper document with the notes in addition to the inserted information 314. In one example, an electronic representation of a user's paper document becomes a template for a media document that includes the inserted information 314. Thus, a user may view the media document and, if more information is desired, the inserted information may be selected and related recorded information 120 may be accessed and viewed.

[0074] Also, composite electronic representation 122 may be a different document than the received electronic representation. The different document may include any or all of the features from the electronic representation received. For example, a user may desire a different format than the electronic representation of the paper document. The slide images or notes in the electronic representation of the paper document may be removed or moved to a different location on the page. For example, composite electronic representation 122 with just the user's notes and inserted information 314 may be generated.

[0075] Composite electronic representation 122 may be stored in various formats. For example, composite electronic representation 122 may be a PDF, HyperText Transfer Language (HTML), Flash, MS Word, etc. formatted document. In one embodiment, the

format supports the insertion of information that may be used to link to recorded information 120.

[0076] As described above, features extracted are used to determine matching information.

In alternative embodiments, other types of information may be used to determine matching

information. For example, bar codes found on a paper document may be used to determine portions of recorded information 120. For example, in one embodiment, documents may be

printed with barcodes associated with each slide. The bar codes may be used to make the link between the slides and the recorded information. For example, techniques described in U.S.

Application No. 10/660,867, entitled "TECHNIQUES FOR ACCESSING INFORMATION

CAPTURED DURING A PRESENTATION USING A PAPER DOCUMENT FOR THE PRESENTATION", filed September 12, 2003 (Attorney Docket No. 15358-009430US).

[0077] In another embodiment, barcodes, or some other markings may be used to represent signature information of each slide. This signature information may include the text from the

slide, image feature vectors, etc. The signature is generated at the time of generating the

document (e.g., during printing). The signature may also include information regarding the

location on the document where the information related to recorded information 120 may be inserted. After the document image is captured (e.g., after scanning), these printed markings

are identified (extracted & decoded) and used for matching, accessing, and inserting the information related to recorded information 120.

[0078] Fig. 4 depicts a possible output after information 314 is selected according to one embodiment of the present invention. Information 314 includes an image but it should be understood that information other than an image may be selected, such as an icons, links, text, pictures, video, audio, etc.

[0079] As shown, an interface 502 may be displayed when an image in information 314 is

selected. Interface 502 includes a window 504 that displays recorded information 120 and a window 506 that includes an image of a slide 304. For discussion purposes, it assumed that a

user has selected an image in information 314-3. After the selection, association information for the image is used to access recorded information 120. For example, the association

information may be source and time information. The source information may used to access

a presentation and the time information is used to determine a portion of the presentation at the time. For example, the time information may be a start time where a slide 304-3 was

displayed.

[0080] Window 504 includes a media player and may be used to display a portion of accessed recorded information 120. As shown, recorded information 120 is displayed in a media player window 508. In this case, a starting image corresponds to the image displayed in information 314-3. A user then may select play in the media player and portions of the presentation are played. Alternatively, the accessed recorded information 120 may be automatically played.

[0081] An image of slide 304-3 may also be displayed in window 506. Also, the original slide in the *.ppt file may be displayed in window 506. Thus, a user may watch a presentation in window 504 in addition to viewing slide 304-3. Accordingly, a user may not need to view the paper copies of the document. Also, window 506 may include other information, such as the notes that a user took as shown in Fig. 3B. Additionally, metadata that is determined may also be displayed in window 506.

[0082] Fig. 5 is a simplified block diagram 600 of modules that may be used to implement an embodiment of the present invention. The modules may be implemented in software or hardware, or combinations thereof. The modules depicted in Fig. 5 are merely illustrative of an embodiment incorporating the present invention and do not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[0083] An electronic representation receiver 602 receives an electronic representation of a paper document. In one embodiment, the electronic representation may be received from a scanner that scanned a paper document to generate the image. Also, an electronic copy of a paper document may be received.

[0084] A feature extractor 604 receives the electronic representation and is configured to extract features from the image. For example, slide images are extracted from the electronic representation of the paper document. As discussed above, various techniques may be used to determine individual slide images.

[0085] A comparer 606 receives the extracted features and is configured to determine matching information for the extracted features. In one embodiment, a database 608 that stores recorded information 120 and association information 609 is queried. The extracted features are compared to recorded information 120 to determine matching information. Association information for the matching information in recorded information 120 may be

determined also. The association information may be used to access portions of presentations.

[0086] An information inserter 610 receives the matching information, association information, and image. Also, recorded information 120 (e.g., audio and video information), metadata, etc. may be received. Information inserter 610 is configured to determine information to insert and to generate a composite electronic representation 122, as described above. For example, information 314 related to the matching information in recorded information 120 is inserted in composite electronic representation 122. Information 314 may also be associated with the association information. In this case, when the inserted information 314 is selected, the association information may be used to access portions of recorded information 120. Also, recorded information 120 (e.g., audio and video information), metadata, etc. may be inserted into electronic representation 310. Thus, database 608 does not need to be accessed when recorded information 120 is played.

Applications:

I. Play program

[0087] Although embodiments of the present invention have been described using presentation recordings, it will be recognized that embodiments described may be used with recorded information other than presentation recordings. A paper document may be various forms including any medium that includes recognizable features you can extract. For example, a paper document may be a program of a play. Before a play in a theater starts, the program of the play is distributed to the audience. The program may include some scenes from the play, images of actors, or some text from the play. The play is then recorded. After the play, composite electronic representation 122 of the play program is received. For example, a user may scan or capture the play program with their digital camera. Using processing described above, composite electronic representation 122 includes information related to the recorded play. The processing may include comparing the play scenes in the program to the captured video frames of the recorded play, comparing the actor's pictures in the program to the face recognition results from the play recording, and comparing the text in the program to the captured audio (speech recognition) to determine matching information.

[0088] Information related to the matching information may then be inserted into composite electronic image. Composite electronic representation 122 of the play program may have objects inserted that are associated with portions of the play. A user may store

composite electronic representation 122 in their digital camera. Alternatively, composite electronic representation 122 may be e-mailed to the user or to others for sharing purposes. Accordingly, when a user is interested in a feature in the program, the inserted information may be selected and a portion of the recorded play may be accessed and played in one
5 example.

II. Symphony

[0089] Another application example is as follows, before a symphony practice takes place, music notes are printed and distributed to the players as paper documents. During the symphony, music is recorded, and players may take notes on the documents. After the
10 practice, electronic representations of the documents are received. An association between the recorded audio is determined by OCR'ing the captured music notes and automatically extracting notes from the audio and matching them. The user then may receive composite electronic representation 122, such as a PDF document, that contains the scanned music notes, the personal notes, and inserted information that associates the information to the audio
15 recorded during the practice (or audio played by another symphony). Composite electronic representation 122 may also be helpful to another person for practicing who may have missed the original practice.

[0090] Techniques for performing segmentation, techniques for matching, and techniques for determining metadata will now be described.

20 Segmentation

[0091] Embodiments of the present invention may segment an electronic representation of a paper document using the following process. Horizontal and vertical projections of an electronic representation are first obtained. It should be noted that some pre-processing of the electronic representation may be required before this step, such as skew correction, down
25 sampling, smearing, connecting component analysis, etc.

[0092] The distance between extracted projections to the projections obtained from a set of document templates is computed. Fig. 6 depicts possible templates that may be included in the set of document templates. Document templates 702 include possible layouts that may have been used to create the pages. The layouts include different images. For example,
30 documents 702-1 and 702-2 include layouts for slide images. Document template 702-1 includes two columns of slide images 704. Document template 702-2 includes a single

column of slide images 706. Document template 702-3 also includes a left column that includes three slide images 708 and a right column that includes three areas 710 where a user may enter notes. The extracted projection is then compared to document templates 702. In one embodiment, these templates are used to create paper documents. For example, the images shown do not have to be images of slides. Rather, windows that indicate a slide should be placed in certain locations where the slide images are shown in Fig. 6 may be used.

[0093] The document template 702 that has a minimum distance to the document image is then determined. For example, a page that includes three slide images in a left column and a space for notes in the right column may substantially match document template 702-3.

[0094] The document image is then segmented into rectangular regions using the slide placement information of a matched document template 702. For example, because the slides are located in the left column with a certain spacing, the document may be segmented into portions that include individual portions that include a slide.

Matching Techniques

[0095] In one embodiment, the following techniques may be used to determine association information using the electronic representation of the document and presentation recording document. A document-matching algorithm receives an image I_i as input and it compares that to the database of recorded Presentation recorder documents. This may be referred to as a presentation-matching step. It locates every presentation recorder document that could have been captured during the presentation where the slides in the document are presented. The next step, called slide matching, maps each slide image in the segmented document image onto the slide images captured by the Presentation Recorder in one presentation session.

[0096] The presentation-matching algorithm applies OCR to the Presentation Recorder document and it saves the text it outputs together with an indication of the page that the text occurred on. An issue is the presence of potentially a large number of both duplicate and spurious images in the Presentation Recorder document caused by people going back and forth in their PowerPoint file while they give a talk, the use of custom animations that cause only minor differences to occur in the captured images, and the use of video clips that can cause the Presentation Recorder to capture hundreds of frames. A pseudo code statement according to one embodiment of the presentation-matching algorithm is shown below.

```

for every slide s in Document Ii
  for every word n-gram w in s
    for every Presentation recorder document Pj containing w
      add Pj to the solution set ss
5      ++score{Pj}
end

```

```

for each Presentation Recorder document Pj in ss
  if (score{Pj}/num_words{Pj} > t1 &&
10  % pages in Pj with > t2 n-grams in Ii > t3)
    then return(Pj);
end

```

[0097] The first step looks up all the Presentation Recorder files that contain each word n-gram in the Document I_i and increments a score for that document. The second step
15 considers the Presentation Recorder files with a minimum percentage of their words in I_i, as specified by t₁, and determines the percentage of pages with more than t₂ of their n-grams in I_i. If this exceeds t₃, then we say this P_j is the matching Presentation Recorder document.

[0098] The presentation-matching algorithm is tolerant of OCR errors in the Presentation Recorder document. Since it contains color jpeg images, we expected the OCR would make
20 many mistakes. However, because presenters typically choose fonts carefully and use short text phrases, errors may not be a significant problem. Also, the requirement that a percentage of the pages in the Presentation Recorder file be contained in the Document takes into account duplicate and spurious images. These factors let thresholds be set liberally.

[0099] An additional consideration in the design of the presentation-matching algorithm
25 was the use of existing full text indexes. This was achieved by using word n-grams as query terms to propose potentially matching PowerPoint files. This is supported by almost every full text index, e.g., Google.

[0100] The slide-matching algorithm determines images in the Document I_i that match each
slide in the Presentation Recorder files located by the presentation-matching algorithm. It
30 uses a combination of string matching on the OCR results for the images that contain text, and edge histogram matching for images that do not contain text. Examples of slide matching techniques are described in more detail in U.S. Application No. 10/412,757,

entitled "AUTOMATED TECHNIQUES FOR COMPARING CONTENTS OF IMAGES",
filed April 11, 2003 (Attorney Docket No. Docket No. 15358-009500US).

Metadata

Text and keywords

5 [0101] Text is extracted from each captured screen image. Text localization and
binarization of each electronic representation is achieved with a scheme that uses not only the
luminance component of a captured image, but also its color components. This is because,
unlike common document images, background/foreground contrast in a slide electronic
representation may be obtained by color contrast as well as luminance contrast. Details of
10 this scheme are described in U.S. Application No. 10/412,757, entitled "AUTOMATED
TECHNIQUES FOR COMPARING CONTENTS OF IMAGES" (Attorney Docket No.
Docket No. 15358-009500US), filed April 11, 2003. A commercial OCR package may be
used to extract text from the binarized text regions. Extracted text is indexed in XML format
with the captured image and line number. Keywords are found by TF-IDF analysis
15 performed on the text extracted from all the screen captured images for a presentation
session.

Electronic representation features

[0102] A number of image feature vectors, i.e., edge histogram and color layout [ID-RII-
311], are computed for the screen capture images. These features are later employed for
20 duplicate slide detection and linking screen images to the original presentation slides.

Symbolic presentation slides

[0103] Presenters can submit the original presentation document to the server prior to their
talk. After the submission, presentation file is assigned an ID, ID_k , and text and slide titles
are extracted from each slide by parsing the presentation file. A rendered JPEG electronic
25 representation of the slide, S_i , is extracted and used to compute the edge histogram and color
layout feature vectors. Presentation file, JPEG images, text, titles, and feature vectors are
indexed and placed to a directory of *unresolved* presentations.

[0104] After each presentation session finishes, image features and text extracted from the
screen capture images are matched with the image features and text extracted from the
30 unresolved presentation files. When there is a match found, the file is removed from the

directory of *unresolved* presentations and linked to the recorded presentation. After presentation level matching, electronic representation and text features are used to match each presentation slide, S_i , to a set of screen capture images, $\{C_1, \dots, C_n\}$. The electronic representation matching process is 98% accurate.

5 **Key frame extraction**

[0105] The conference room is equipped with 2 cameras, a Pan-Zoom-Tilt camera and an omni-directional camera with 360° capture. The PTZ camera focuses on either the presenter or the entire conference room. PTZ camera location is controlled by the meeting room portal and each time the camera location changes a keyframe is extracted from the video sequence.

10 The omni-directional camera is placed in the middle of the conference room and captures a panoramic view of the room. Four microphones are attached to the camera and sound source localization (SSL) is performed real-time on the 4 channel audio. Each time the sound source changes direction, a key frame showing only a perspective view from the direction of the sound is extracted from the panoramic video. Fig. 7 shows these key frames 802, which are

15 very useful when navigating meeting/presentation recordings as they coincide with speaker changes. All the extracted key frames are indexed with their source and timestamps.

Time spent on each slide

[0106] The amount of time spend on a presentation slide, sT_i , can be a good indicator of the

importance of that particular slide and computed as

$$sT_i = \frac{1}{pT} \sum_{C_n \approx S_i} T(C_{n+1}) - T(C_n), \text{ where}$$

20 pT is the total presentation time, S_i is the presentation slide, C_n is the n^{th} captured screen electronic representation matching S_i , $T(C_n)$ is the timestamp of C_n in seconds.

Question & answer activity

25 [0107] The amount of discussions, questions, comments around a particular presentation slide can be an indication of the interest around the topic it discusses. On the other hand, segmentation of speakers, Q&A sessions, are quite challenging and often require prior training of the speaker segmentation systems. Our approach for identifying audio segments with Q&A activity is somewhat different. It is based ion SSL and works quite robustly in

30 practice. Our experiments showed that during a presentation, 93% of the time the speaker stays in the same 20 degree azimuth(region) of the SSL device. Clearly, the azimuth range

could change based on the conference room setup. Nevertheless, for most conference room setups, it is often a reasonable assumption to make that the presenter has a limited platform to move around. Let's indicate this range of this platform wrt the SSL with $[\alpha_{s1} \alpha_{s2}]$. Assuming that there is no audience between the presenter and the SSL device, sound coming from the direction other than the presenter can be interpreted as a comment or a question from an audience member. The question and answer activity for a given presentation slide S_i is defined as the number of times the sound source change directions between the audience and the presenter as follows:

$$sQA_i = \frac{1}{pD} \sum_{C_n \approx S_i} \sum_{t=T(C_n)}^{T(C_{n+1})-1} |D(t) - D(t+1)|, \text{ where } C_n \text{ is the } n^{\text{th}} \text{ captured screen image}$$

matching S_i , $T(C_n)$ is the timestamp of C_n in seconds, pD is the QA activity for the whole

$$pD = \sum_{t=0}^{pT} |D(t) - D(t+1)|, \text{ presentation, i.e., } pT \text{ is the total presentation time, and } D(t)$$

function is defined as $D(t) = \begin{cases} \alpha_{s1} \leq SSL(t) \leq \alpha_{s2} & 1 \\ \text{ow} & 0 \end{cases}$, where $SSL(t)$ is the azimuth of sound direction at t .

Notes activity

The amount of note taking activity takes place around certain topics in a seminar is most of the time directly relevant to the interest of the audience to that topic. A notes activity measure is computed as follows:

$$sN_i = \frac{1}{\eta(0, pT)} \sum_{C_n \approx S_i} \eta(T(C_n), T(C_{n+1}))$$

, where $\eta(t_1, t_2)$ returns the number of note entries associated with the $[t_1 \ t_2]$ timeframe. It is important to note that the electronic note taking interface allows the association of notes with the current time as well as with the past times and the η function returns the note entries based on the associated time rather than the entry time.

[0108] Fig. 8 is a simplified block diagram of data processing system 900 that may be used to perform processing according to an embodiment of the present invention. As shown in

Fig. 8, data processing system 900 includes at least one processor 902, which communicates with a number of peripheral devices via a bus subsystem 904. These peripheral devices may include a storage subsystem 906, comprising a memory subsystem 908 and a file storage subsystem 910, user interface input devices 912, user interface output devices 914, and a network interface subsystem 916. The input and output devices allow user interaction with data processing system 900.

[0109] Network interface subsystem 916 provides an interface to other computer systems, networks, and storage resources. The networks may include the Internet, a local area network (LAN), a wide area network (WAN), a wireless network, an intranet, a private network, a public network, a switched network, or any other suitable communication network. Network interface subsystem 916 serves as an interface for receiving data from other sources and for transmitting data to other sources from data processing system 900. For example, data processing system 900 may access stored recorded information for a presentation and XML data structures via network interface subsystem 916. Embodiments of network interface subsystem 916 include an Ethernet card, a modem (telephone, satellite, cable, ISDN, etc.), (asynchronous) digital subscriber line (DSL) units, and the like.

[0110] User interface input devices 912 may include a keyboard, pointing devices such as a mouse, trackball, touchpad, or graphics tablet, a scanner, a barcode scanner, a touchscreen incorporated into the display, audio input devices such as voice recognition systems, microphones, and other types of input devices. In general, use of the term "input device" is intended to include all possible types of devices and ways to input information to data processing system 900.

[0111] User interface output devices 914 may include a display subsystem, a printer, a fax machine, or non-visual displays such as audio output devices. The display subsystem may be a cathode ray tube (CRT), a flat-panel device such as a liquid crystal display (LCD), or a projection device. In general, use of the term "output device" is intended to include all possible types of devices and ways to output information from data processing system 900.

[0112] Storage subsystem 906 may be configured to store the basic programming and data constructs that provide the functionality of the present invention. For example, according to an embodiment of the present invention, software modules implementing the functionality of the present invention may be stored in storage subsystem 906. These software modules may be executed by processor(s) 902. Storage subsystem 906 may also provide a repository for

storing data used in accordance with the present invention. Storage subsystem 906 may comprise memory subsystem 908 and file/disk storage subsystem 910.

[0113] Memory subsystem 908 may include a number of memories including a main random access memory (RAM) 918 for storage of instructions and data during program execution and a read only memory (ROM) 920 in which fixed instructions are stored. File storage subsystem 910 provides persistent (non-volatile) storage for program and data files, and may include a hard disk drive, a floppy disk drive along with associated removable media, a Compact Disk Read Only Memory (CD-ROM) drive, an optical drive, removable media cartridges, and other like storage media.

[0114] Bus subsystem 904 provides a mechanism for letting the various components and subsystems of data processing system 902 communicate with each other as intended. Although bus subsystem 904 is shown schematically as a single bus, alternative embodiments of the bus subsystem may utilize multiple busses.

[0115] Data processing system 900 can be of varying types including a personal computer, a portable computer, a workstation, a network computer, a mainframe, a kiosk, or any other data processing system. Due to the ever-changing nature of computers and networks, the description of data processing system 900 depicted in Fig. 8 is intended only as a specific example for purposes of illustrating the preferred embodiment of the computer system. Many other configurations having more or fewer components than the system depicted in Fig. 8 are possible.

[0116] Although specific embodiments of the invention have been described, various modifications, alterations, alternative constructions, and equivalents are also encompassed within the scope of the invention. The described invention is not restricted to operation within certain specific data processing environments, but is free to operate within a plurality of data processing environments. Additionally, although the present invention has been described using a particular series of transactions and steps, it should be apparent to those skilled in the art that the scope of the present invention is not limited to the described series of transactions and steps. It should be understood that the equations described above are only illustrative of an embodiment of the present invention and can vary in alternative embodiments of the present invention.

[0117] Further, while the present invention has been described using a particular combination of hardware and software, it should be recognized that other combinations of

hardware and software are also within the scope of the present invention. The present invention may be implemented only in hardware, or only in software, or using combinations thereof.

[0118] The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense. It will, however, be evident that additions, subtractions, deletions, and other modifications and changes may be made thereunto without departing from the broader spirit and scope of the invention as set forth in the claims.

5